On the Entropy Region of Discrete and Continuous Random Variables and Network Information Theory

Sormeh Shadbakht and Babak Hassibi
Electrical Engineering Department
California Institute of Technology
Pasadena, California 91125
Email: sormeh, hassibi@caltech.edu

***Abstract*— We show that a large class of network information theory problems can be cast as convex optimization over the convex space of entropy vectors. A vector in $2^n - 1$ dimensional space is called entropic if each of its entries can be regarded as the joint entropy of a particular subset of n random variables (note that any set of size n has $2^n - 1$ nonempty subsets.) While an explicit characterization of the space of entropy vectors is well-known for $n = 2, 3$ random variables, it is unknown for $n > 3$ (which is why most network information theory problems are open.) We will construct inner bounds to the space of entropic vectors using tools such as quasi-uniform distributions, lattices, and Cayley's hyperdeterminant.**

I. Introduction

The growing interest in information transmission over networks has led to the study of limits of information transfer among many users, i.e., multiuser information theory. However in contrast to the single user case which is a well-understood problem, the area of multiuser information theory has many long-standing open problems. Determining the capacity region of information networks in general requires one to solve an infinite-letter non-convex optimization problem which is an extremely difficult task. Therefore the capacity regions of even simple networks such as the relay or interference channel are still unknown.

Let a network be represented by a directed acyclic graph $\mathcal{G} = \{V, E\}$ where the nodes denote either a channel or a network operation (shown by circles and squares respectively) and each edge corresponds to a signal. Source messages and destination signals are denoted by s and d and each destination requires a subset of the source messages as its demand. Using this model, the point-to-point communication network is depicted in Fig. 1. The capacity of this memoryless network is clearly,

$$C = \max_{p_S(\cdot)} I(S;D) = \max_{p_S(\cdot)} \{H(D) - H(D|S)\}, \qquad (1)$$

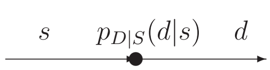


Fig. 1. A point-to-point communication problem.

This work was supported in part by the National Science Foundation through grant CCF-0729203, by the David and Lucille Packard Foundation, by the Office of Naval Research through a MURI under contract no. N00014-08-1-0747, and by Caltech's Lee Center for Advanced Networking.

where $p_S(\cdot)$ is the input distribution and $H(D)$ and $H(D|S) = H(D,S) - H(S)$ are the usual entropy and conditional entropies. Note that this is a convex optimization problem since $I(S;D)$ is a concave function of the input distribution for a given $p_{D|S}(d|s)$. Moreover, since the entropies are over a single channel use, (1) is referred to as single letter.

Now consider a general acyclic discrete memoryless network (Fig. 2) where every source s_i wants to transmit data to its corresponding receiver d_i at a rate R_i. Note that without loss of generality we can assume that the number of sources and destinations are equal. This can be achieved by repeating the sources and destinations as many times as necessary.

The rate region for reliable communication is known to be obtained from (see e.g. [1], [2], [3]):

$$\mathcal{R} = \text{cl}\left\{ R_i \mid R_i < \frac{1}{T}\left(H(D_i^T) - H(D_i^T|S_i^T)\right)\right\} \quad \text{as } T \to \infty, \qquad (2)$$

where S_i^T and D_i^T are the vectors of source and channel random variables over T channel uses and $\text{cl}\{\cdot\}$ refers to the closure of the set as $T \to \infty$. Although this characterization might not be surprising, computing it is a difficult problem. Equivalently, characterizing $\mathcal{R}$ can be done through tangent hyperplanes by solving the following optimization problem:

$$\lim_{T\to\infty} \sup_{\substack{p_{S_i^T}(\cdot) \text{ and} \\ \text{network operations}}} \sum_{i=1}^{m} \alpha_i \frac{1}{T}\left(H(D_i^T) - H(D_i^T|S_i^T)\right), \qquad (3)$$

where $\{\alpha_i\}_{i=1}^m$ is the normal vector to the tangent hyperplane, and "network operations" represents all permissible internal operations of the network.

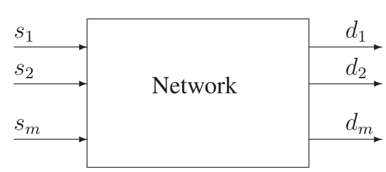


Fig. 2. A communication problem over an acyclic memoryless network.

This problem is extremely difficult since it is both infinite letter (the number of the channel uses is going to infinity) and non-convex in the source distributions and network operations.

In this paper we tackle the network information theory problem from the entropy vectors perspective. This viewpoint resolves the infinite-letter and non-convexity issues and reveals the heart of network information theory problems: *"the entropy region characterization"*. Although complete characterization of this region is still a formidable task in general –accounting for many network information theory problems to be open– partial characterizations can be used through this framework to yield tighter bounds on the capacity of information networks.

The rest of the paper is organized as follows. In the next section it is shown that the network information theory problem can be cast as an optimization problem over the convex set of channel-constrained entropic vectors. This unveils the importance of characterizing the entropy region in network information theory. In section III properties of the entropy region are studied and two approaches towards characterizing this region namely, lattice structure and Gaussian distributed random variables, are presented.

II. Network Problem and Entropy Vectors

Due to the difficulty of obtaining the rate region for a network problem via infinite-letter characterization, (2) or (3) have rarely been used in the literature [4]. It turns out that with a slightly different definition of entropy, the non-convexity and infinite-letter issues can be resolved and (3) can be cast as a convex optimization problem.

Let $\mathcal{N} = \{1, \cdots, n\}$ and consider n discrete random variables $X_i,\ i \in \mathcal{N}$ with alphabet size N. For any $\alpha \subseteq \mathcal{N}$, let $h(X_\alpha)$ be the joint entropy of random variables $X_i,\ i \in \alpha$ normalized by the log of the alphabet size. The $2^n - 1$ dimensional vector whose components are the normalized joint entropies $h(X_\alpha)$, is called the *normalized entropy vector* of $X_1, \cdots, X_n$. Conversely any $2^n - 1$ dimensional vector that can be considered as the normalized entropy vector of n random variables of some alphabet size N is called *normalized-entropic*. The space of all normalized-entropic vectors of dimension $2^n - 1$ is denoted by Ω_n^*.

In (3), the alphabet size of the vector-valued source or destination random variables over T channel uses is equal to N^T and therefore the $\frac{1}{T}H(.)$ terms are proportional to the normalized entropies. Furthermore, considering the normalized entropy makes the entropy region finite, i.e,

$$h(X_\alpha) \leq |\alpha|, \tag{4}$$

and provides an easy proof for the convexity of the closure of the entropy region. Therefore we believe that our definition which is based on considering normalized entropies is more natural than the conventional non-normalized form (see e.g. [5].) Γ_n^* denotes the space of non-normalized entropies.

Theorem 1 (Convexity of $\bar{\Omega}_n^*$): The closure of the set of entropic vectors, $\bar{\Omega}_n^*$ is convex [6].

Proof 1: [Time Sharing] Consider two sets of random variables $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ with alphabet sizes N_x and N_y respectively, and let $h_x, h_y \in \Omega_n^*$ be the corresponding normalized entropy vectors. Make n_x independent copies of the first set and n_y independent copies of the second set to obtain a new set of random variables with alphabet size $N_x^{n_x} N_y^{n_y}$. The resulting normalized entropy vector is then,

$$\frac{n_x \log N_x}{n_x \log N_x + n_y \log N_y} h_x + \frac{n_y \log N_y}{n_x \log N_x + n_y \log N_y} h_y, \tag{5}$$

which since n_x and n_y are arbitrary, proves the convexity of the closure of Ω_n^*.

Proof 2: [Convex Combination of Distributions] Consider the convex combination of the distributions of variables $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$ with alphabet size N. Make T independent copies of each of the sets of random variables X_i and Y_i and consider the distribution

$$\begin{aligned} p_{Z_1,\ldots,Z_n}(z_1^T, \ldots, z_n^T) = p_\theta \prod_{t=1}^{T} p_{X_1,\ldots,X_n}(z_1^t, \ldots, z_n^t) + \\ (1 - p_\theta) \prod_{t=1}^{T} p_{Y_1,\ldots,Y_n}(z_1^t, \ldots, z_n^t). \end{aligned} \tag{6}$$

Now for any $\mathcal{S} \subseteq \{1, \ldots, n\}$, we have

$$\underbrace{H(Z_\mathcal{S}^T|\theta)}_{=p_\theta H(X_\mathcal{S}^T) + (1-p_\theta)H(Y_\mathcal{S}^T)} \leq H(Z_\mathcal{S}^T) \leq \underbrace{H(Z_\mathcal{S}^T, \theta)}_{=H(Z_\mathcal{S}^T|\theta) + H(p_\theta)} . \tag{7}$$

Normalizing by $\log N^T$ yields

$$\begin{aligned} p_\theta h_x + (1 - p_\theta) h_y \leq h_z \leq p_\theta h_x + (1 - p_\theta) h_y \\ + \frac{-p_\theta \log p_\theta - (1 - p_\theta)\log(1 - p_\theta)}{T \log N}, \end{aligned} \tag{8}$$

which shows the convexity of the closure as $T \to \infty$. ■

Let us now return to the network problem (3) and study it from the perspective of entropy vectors. Let $X_1, \cdots, X_n$ represent all the random variables of the network including all the sources, destinations, as well as any internal random variables and define h as the normalized entropy vector of these variables. As a result, the objective function of the problem (3) can be written as a linear combination of the joint entropies of the network variables allowing us to reformulate (3) as,

$$\sup \alpha^T h, \tag{9}$$

where α is the vector of coefficients and $(\cdot)^T$ refers to transpose. This optimization is subject to $h \in \bar{\Omega}_n^*$ and subject to the constraints imposed by the network. These constraints are of two types:

1) **Topological constraints:** Topological constraints have to do with the information flow and causality conditions in the network and they introduce linear constraints on the entries of the entropy vector. In a ("non-source") operational node, output variables are determined by the inputs, i.e. if $\{X_{i_p}\}_{p=1}^k$ are the incoming messages and $\{X_{j_q}\}_{q=1}^l$ are the outgoing messages (Fig. 3), then we have the following linear constraint on the entropy,

$$h\left(X_{j_q} | X_{i_1}, \ldots, X_{i_k}\right) = 0, \tag{10}$$

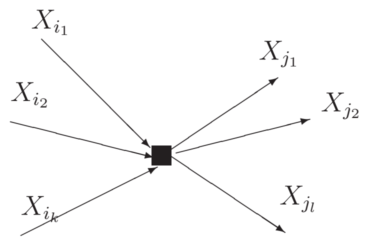


Fig. 3. Topological constraints at any operational node.

or equivalently:

$$h\left(X_{j_q}, X_{i_1}, \ldots, X_{i_k}\right) - h\left(X_{i_1}, \ldots, X_{i_k}\right) = 0, \qquad (11)$$

for all $q = 1, \ldots l$. At sources we have $h(S_i, S_j) - h(S_i) - h(S_j) = 0$, if sources i and j are independent or $h(S_i, S_j) = h(S_i) = h(S_j)$, if sources i and j are identical.

2) **Channel constraints:** Channel constraints are due to the channel conditional probability distributions. However, as opposed to the topological constraints, they do not directly translate into entropies.

Consider an internal channel of a network with input X_i and output X_j (Fig. 4). Then the probability distributions will be constrained as follows:

$$p(X_i, X_j) = p(X_j|X_i)p(X_i), \qquad (12)$$

or, equivalently,

$$\begin{aligned} &\int \prod_{k \neq i,j} dX_k \; p(X_1, \ldots, X_n) \\ &= p(X_j|X_i) \int \prod_{k \neq i} dX_k \; p(X_1, \ldots, X_n), \qquad (13) \end{aligned}$$

which is a linear constraint on the joint probability distribution. Note that the validity of the proofs of Theorem 1 will not be affected when the underlying distributions satisfy some linear channel constraints. Therefore if we denote the space of entropic vectors that are constrained by the discrete memoryless channels in the network by $\Omega^*_{n,c}$, we have [6],

Theorem 2 (Channel-Constrained Entropic Vectors): Closure of the channel constrained entropic vectors, $\bar{\Omega}^*_{n,c}$, is convex.

We can now state the network problem of (3) as a convex optimization [6],

Theorem 3 (Network Problem as a Convex Optimization): The network information theory problem can be cast as the following optimization problem,

$$\max_{h \in \bar{\Omega}^*_{n,c}, Ah=0} \alpha^T h, \qquad (14)$$

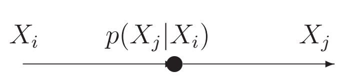


Fig. 4. A channel internal to the network.

where $\bar{\Omega}^*_{n,c}$ denotes the convex space of channel-constrained entropic vectors and $Ah = 0$ represents the topological constraints. [1]

Note that this formulation is significant in two ways: We have circumvented the "infinite-letter characterization" and have also convexified the problem. It also reveals the importance of characterizing the entropy region in solving the network information theory problems.

Although characterizing $\Omega^*_{n,C}$ is still a very difficult task (as we will see in the next section), it turns out that even without complete characterization we can get some interesting results.

A. Duality and Cutset Bound

Once again consider problem (14). Using the Lagrange multipliers we can write,

$$\max_{h \in \bar{\Omega}^*_{n,c}, Ah=0} \alpha^T h = \max_{h \in \bar{\Omega}^*_{n,c}} \min_{\lambda} \left(\alpha^T h + \lambda^T A h\right). \qquad (15)$$

Since the problem is convex, we can interchange the $\max$ and $\min$ via the duality argument to obtain,

$$\max_{h \in \bar{\Omega}^*_{n,c}, Ah=0} \alpha^T h = \min_{\lambda} \max_{h \in \bar{\Omega}^*_{n,c}} \left(\alpha^T h + \lambda^T A h\right). \qquad (16)$$

Any choice of λ yields an upper bound for the original problem. We partition the nodes of the network into two sets, one containing all the sources and the other all the destinations. Any node whose edges do not cross the cut, we set its corresponding λ to zero. This makes the network essentially like a point-to-point problem, since the nodes on each side of the cut can now fully co-operate with each other. As a result, optimizing over the remaining Lagrange multipliers yields the cut capacity.

Therefore we have obtained the cutset upperbound via a duality argument. This resembles the flow networks where the duality between the max-flow and min-cut is well-known [7], [8].

III. Entropy Region

In the networks for which separation of channel and network coding holds ([9], [6]), channels affect the rate region only through their capacities and therefore the optimization over $\Omega^*_{n,C}$ in problem (14) can be replaced by an optimization over the unconstrained entropy region, Ω^*_n, and the set of capacity constraints. Therefore in the remaining we will focus on characterizing the unconstrained entropy region Ω^*_n.

A. What is Known about the Entropy Region?

It is known that the entropy has the following properties,

1) $H(\emptyset) = 0$,
2) For $\alpha \subseteq \beta$: $H(\alpha) \leq H(\beta)$,
3) For any α, β: $H(\alpha \cup \beta) + H(\alpha \cap \beta) \leq H(\alpha) + H(\beta)$,

where $\alpha, \beta \subseteq \mathcal{N} = \{1, \cdots, n\}$. The last property is called the *submodularity property* and all of the above properties follow from the non-negativity of the conditional mutual information. Any information inequality that can be expressed as a positive

[1] Since the constraint set is closed, we can use $\max$, rather than $\sup$.

linear combination of conditional mutual information terms, i.e.,

$$\sum \alpha_i I(A_i; B_i | C_i) \geq 0 \quad \alpha_i \geq 0,$$

is called a *Shannon type inequality*. The space of all $2^n - 1$ dimensional vectors whose entries only satisfy the Shannon inequalities is referred to as $\Omega_n(\Gamma_n)$ as opposed to $\Omega_n^*(\Gamma_n^*)$. The entropy region for two and three random variables has completely been characterized [10], [11],

$$\Omega_2^* \ (\Gamma_2^*) = \Omega_2 \ (\Gamma_2) \quad \text{and} \quad \bar{\Omega}_3^* \ (\bar{\Gamma}_3^*) = \Omega_3 \ (\Gamma_3),$$

where $\bar{\Omega}_3^*(\bar{\Gamma}_3^*)$ refers to the closure of $\Omega_3^*(\Gamma_3^*)$. However for $n = 4$ some *non-Shannon type* information inequalities were found [12] which proved that the entropy region for $n \geq 4$ is a proper subset of $\Omega_n(\Gamma_n)$, i.e.,

$$\Omega_4^* \ (\Gamma_4^*) \subset \Omega_4 \ (\Gamma_4).$$

The non-Shannon inequalities provide outer bounds for the entropy region and it has been shown that no finite number of these inequalities can characterize the entropy region completely [13]. Innerbounds are less often studied in the literature and the most well-known inner region is defined through the *Ingleton inequality* which was first obtained for the ranks of vectors spaces [14]. Let $v_1, \cdots, v_n$ be n vector subspaces and r_α be defined as the rank of the subspace $\oplus_{i \in \alpha} v_i$. Then for any $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \subseteq \{1, \cdots, n\}$ the following inequality holds,

$$\begin{aligned} & r_{\alpha_1} + r_{\alpha_2} + r_{\alpha_1 \cup \alpha_2 \cup \alpha_3} + r_{\alpha_1 \cup \alpha_2 \cup \alpha_4} + r_{\alpha_3 \cup \alpha_4} \\ & - r_{\alpha_1 \cup \alpha_2} - r_{\alpha_1 \cup \alpha_3} - r_{\alpha_1 \cup \alpha_4} - r_{\alpha_2 \cup \alpha_3} - r_{\alpha_2 \cup \alpha_4} \leq 0. \end{aligned} \quad (17)$$

Ingleton bound for entropies can be obtained by replacing r_α by h_α in (17). Although not all entropy vectors satisfy this inequality, entropies of some important classes of distributions satisfy this bound [15].

One way to characterize the entropy region is through determining its tangent hyperplanes by essentially solving the following optimization problem,

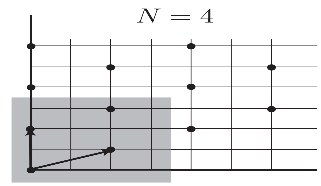
$$\min_{H \in \Gamma_n^*} \sum_{\alpha \subseteq \mathcal{N}} a_\alpha H_\alpha. \quad (18)$$

The arbitrariness in the alphabet size of the random variables makes this optimization difficult. However if we restrict the alphabet size to N and optimize over the joint distribution $p_{X_\mathcal{N}}(x_\mathcal{N})$, we find that the KKT conditions enforce the following condition,

$$\sum_{\alpha \subseteq \mathcal{N}} a_\alpha \log \frac{1}{p_{X_\alpha}(x_\alpha)} = c \quad \text{if} \quad p_{X_\mathcal{N}}(x_\mathcal{N}) \neq 0. \quad (19)$$

Of course there can be many solutions to the above equation. However one solution which does not depend on the point x_α is the one that all its marginals take on either a constant or zero value,

$$p_{X_\alpha}(x_\alpha) = c_\alpha \quad \text{or} \quad 0, \quad (20)$$



$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

Fig. 5. An example of a lattice.

where c_α is a constant independent of x_α. These distributions are referred to as *quasi-uniform*. Computing the entropy of a quasi-uniform distribution is straightforward,

$$H(\alpha) = \log \frac{1}{c_\alpha}. \quad (21)$$

The interesting result of [16] and [17] shows that these distributions are sufficient to characterize the whole entropy region. Let Λ_n denote the space of entropy vectors obtained from quasi-uniform distributions [16],

Theorem 4 (Quasi-Uniform Distributions): $\overline{\text{con}}(\Lambda_n) = \bar{\Gamma}_n^*$, i.e., the convex closure of Λ_n is the closure of Γ_n^*.

To obtain an innerbound for the entropy region of discrete random variables we will henceforth focus on quasi-uniform distributions.

B. *Entropy Region - Discrete Random Variables*

Although quasi-uniform distributions are sufficient to characterize the entropy region, determining all of such distributions is a very hard combinatorial problem. In order to find an inner bound for the entropy region that can be generalized to any number of variables we have to enforce some structure on the distribution. We have considered the lattice structure.
In this model the lattice points are denoted by n-dimensional vectors x, such that each x represents one realization of the n random variables. To be more clear,

$$x = Mz, \quad (22)$$

where $x \in \mathcal{Z}^n$ represents a point of the lattice, $M \in \mathcal{Z}^{n \times n}$ is called the *lattice-generating matrix* and $z \in \mathcal{Z}^n$ is an integer vector. Fig. 5 shows an example of a lattice. We define a probability distribution on this lattice as follows [18],

Definition 1 (Lattice-Generated Distribution): A probability distribution over n random variables, each with alphabet size N, is called *lattice-generated*, if for some lattice $\mathcal{L}(M)$,

$$p_{X_\mathcal{N}}(x_\mathcal{N}) = \begin{cases} c & \text{if } x_\mathcal{N} \in \{0, \ldots, N-1\}^n \cap \mathcal{L}(M), \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

Enforcing the quasi-uniformity on the defined distribution gives the following Lemma [18],

Lemma 1 (Lattice-Generated Quasi-Uniform Distributions): A lattice-generated distribution is quasi-uniform if the lattice has a period that divides N. The latter is true if, and only if, the matrix $M^{-1}N$ has integer entries.

Using (21), computing the corresponding entropy vector of a quasi-uniform lattice-generated distribution becomes simple.

Obtaining the convex hull of the lattice-generated entropies yields the following important results [18],

- The convex hull of the lattice-generated entropy region is tight for $n = 2, 3$ random variables.
- For $n \geq 4$ it gives a polytope inner-bound for the entropy region allowing to solve network problems via a linear program. The inner-bound is at least as tight as the Ingleton inner-region.
- It includes all the scalar and vector linear codes.

C. Entropy Region - Continuous Random Variables

It has been shown that there is a close connection between the entropy region of continuous and discrete variables [19],

Theorem 5 (Discrete and Continuous Entropies):

- A linear continuous information inequality $\sum_\alpha a_\alpha h_\alpha \geq 0$ is valid if and only if its discrete counterpart is valid and for all $i \in \mathcal{N}$, $\sum_{\alpha: i\in\alpha} a_\alpha = 0$.
- A linear discrete information inequality $\sum_\alpha a_\alpha H_\alpha \geq 0$ is valid if and only if it can be written as $\sum_\alpha \beta_\alpha H_\alpha + \sum_{i=1}^n r_i(H_{i,i^c} - H_{i^c}) \geq 0$ for some $r_i \geq 0$, where $\sum_\alpha \beta_\alpha h_\alpha \geq 0$ is a valid continuous information inequality (i^c denotes the complement of i in $\mathcal{N}$).

Therefore one region can be obtained from the other and vice versa. However studying all the possible probability density functions in order to characterize the entropy region seems almost impossible. A natural class of continuous random variables to study first is the class of Gaussian distributed random variables. In fact they are shown to have some properties that makes them desirable to study. The space of normalized Gaussian entropic vectors is a convex cone (via simple concatenation of random variables) and they easily violate the Ingleton bound. Moreover, some non-Shannon type inequalities are tight for Gaussians [20].

Consider n vector-valued[2] Gaussian random variables, $X_1, \cdots, X_n$ with zero mean and covariance matrix R such that X_i is a T-dimensional vector. For any $\alpha \subseteq \{1, \cdots, n\}$, the differential entropy of X_α is obtained from,

$$h_\alpha = \frac{1}{T} \cdot \frac{1}{2} \log \left((2\pi e)^{T|\alpha|} \det R_\alpha \right), \tag{24}$$

where R_α is the $T|\alpha| \times T|\alpha|$ principal minor determined by the rows and columns belonging to the set α. Note that we have normalized the entropy by the dimension of the vector-valued random variables, i.e. T. From (24) it is obvious that studying the entropy of Gaussian random variables requires the study of determinantal inequalities and the relations between the principal minors of a positive definite symmetric matrix [21]. One of the important relations as it turns out is the so-called Cayley's hyperdeterminant [22]. For $n = 2, 3$ we have interestingly obtained that scalar Gaussians can generate the whole entropy region of 2 and 3 continuous random variables [23],

[2]We consider vector-valued random variables since the covariance matrix of n Gaussian variables has only $\frac{n^2+n}{2}$ free parameters while the entropy region is $2^n - 1$ dimensional.

Theorem 6 (Region of 3 Scalar Gaussian Variables): For $n = 3$, the *convex cone* generated by the normalized entropy vectors of three *scalar-valued* jointly Gaussian random variables is the same as the cone generated by entropic vectors.

For $n \geq 4$ this characterization is under investigation where a closely related problem in characterizing the entropy region is the study of necessary and sufficient conditions for a $2^n - 1$ dimensional vector to correspond to all of the principal minors of a symmetric matrix [22], [24].

References

[1] C.E. Shannon, "Two-way communications channels," in *Proc. of the 4th Berkeley Symp. on Mathematical Staistics and Probabililty*. University of California Press, 1961, pp. 611–644.

[2] R. Ahlswede, "A survey of multi-way channels in information theory: 1961-1976," *IEEE Tran. on Inf. Theory*, vol. 23, pp. 1–37, January 1977.

[3] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Tran. on Inf. Theory*, vol. 49, pp. 4–21, January 2003.

[4] R.S. Cheng and S. Verdu, "On limiting characterizations of memoryless multiuser capacity regions," *IEEE Tran. on Inf. Theory*, vol. 39, pp. 609–612, March 1993.

[5] R.W. Yeung, *A first course in information theory*, Kluwer, 2002.

[6] B. Hassibi and S. Shadbakht, "Normalized entropy vectors, network information theory and convex optimization," in *Inf. theory workshop, Bergen, Norway*, 2007.

[7] P. Elia, A. Feinstein, and C.E. Shannon, "Note on maximum flow through a network," *IRE Tran. on Inf. Theory*, vol. 2, pp. 117–119, 1956.

[8] L.R. Ford and D.R. Fulkerson, "Maximal flow through a network," *Canadien Journal of Mathematics*, vol. 8, pp. 399–404, 1956.

[9] R. Koetter, "Separation for source, channel and network coding," in *IEEE Inf. Theory Workshop—Keynote Speech*, 2006.

[10] S. Fujishije, "Polymatroidal dependence structure of a set of random variables," *Information and Control*, vol. 39, pp. 55–72, 1978.

[11] T. S. Han, "A uniqueness of Shannon's information distance and related nonnegativity problems," *J. Comb.,Inform. Syst. Sci.*, vol. 6, no. 4, pp. 320–331, 1981.

[12] Z. Zhang and R. Yeung, "On characterization of entropy function via information inequalities," *IEEE Tran. on Inf. Theory*, vol. 44, no. 4, pp. 1440–1452, 1998.

[13] F. Matus, "Infinitely many information inequalities," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2007, pp. 41–44.

[14] A. W. Ingleton, "Representation of matroids," in *Combinatorial mathematics and its applications, D. Welsh, Ed. London: Academic Press*, 1971, pp. 149–167.

[15] T. H. Chan, "Capacity region for linear and Abelian network codes," in *Information theory and applications workshop, San Diego, CA*, 2007.

[16] T.H. Chan, "A combinatorial approach to information inequalities," *Communications In Inf. and Systems*, vol. 1, no. 3, pp. 241–254, 2001.

[17] T.H. Chan and R.W. Yeung, "On a relation between information inequalities and group theory," *IEEE Tran. on Inf. Theory*, vol. 48, no. 7, pp. 1992–1995, 2002.

[18] B. Hassibi and S. Shadbakht, "On a construction of entropic vectors using lattice-generated distributions," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2007, pp. 501–505.

[19] T. H. Chan, "Balanced information inequalities," *IEEE Tran. on Inf. Theory*, vol. 49, no. 12, pp. 3261–3267, 2003.

[20] Radim Lnenicka, "On the tightness of the Zhang-Yeung inequality for Gaussian vectors," *Communications in information and systems*, vol. 3, no. 1, pp. 41–46, 2003.

[21] Charles R. Johnson and Wayne W. Barrett, "Determinantal inequalities for positive definite matrices," *Discrete Mathematics*, vol. 119, pp. 97–106, 1993.

[22] O. Holtz and B. Sturmfels, "Hyperdeterminantal relations among symmetric principal minors," *J. of Algebra*, vol. 316, pp. 634–648, 2007.

[23] B. Hassibi and S. Shadbakht, "The entropy region for three Gaussian random variables," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2008.

[24] S. Shadbakht and B. Hassibi, "Cayleys hyperdeterminant, the principal minors of a symmetric matrix and the entropy region of 4 Gaussian random variables," in *Forty-Sixth Allerton Conference*, 2008.